

Remarks

Reconsideration of the application is respectfully requested in view of the foregoing amendments and following remarks. With entry of amendments included herein, claims 1-2 and 4-37 are pending in this application. Claims 1, 9, 23, 31, 36, and 37 are independent. No claims have been allowed. Claims 3, 10 and 17 have been previously canceled without prejudice. Claim 37 is rejected under 35 U.S.C. 102(b) as being anticipated by USP 5,729,471 to *Jain* et al. ("*Jain*"). Claims 1-2, 4-9, 11-16, and 18-36 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Jain* in view of USP 5,612,743 to *Lee* ("*Lee*").

1. Cited References

In the interest of shared understanding, the following descriptions of the cited references are given.

A. *Jain*

Jain describes a system that can track specific football players in real time using fixed cameras whose location and movement is known. [See *Jain* 19:34-42.] However, only a "rudimentary, prototype, MPI video system" is described. [*Jain*, 24:22.] In this prototype, three cameras are used to film 10 seconds of a football game. The camera information is then processed using both the video and information known about the camera (camera position, camera rotation angle) to track a specific player. [See *Jain* 23:42-44.] The tracking has two components. In one, a three-dimensional cursor is placed on each of the three video sequences roughly in the position of the tracked player, the cursor pointing in the direction the player is moving. [See *Jain* 25:4-6.] FIGS 10a-10c.] In the other, the camera with the best view of the player can be chosen. [*Jain* 24:5-11.]

To process the video, rudimentary three-dimensional analysis is performed separately on each camera's video image. [See *Jain* 23:61-63.] Specifically, for each video sequence, key frames are selected manually for every thirty frames. [*Jain* 23:64-65.] Within the selected

key frames, a person manually selects feature points. The feet and identity of each player are marked, along with the goal marks visible in the frame. [Jain 22:11-15.] Then, the locations of the marked players and goal marks within successive key frames are interpolated—essentially discarding the other 29 frames, as shown by the following passage: “For frames in between, player position and camera status is estimated by interpolation between key frames by proceeding under the assumption that coordinate values change linearly between a consecutive two key frames.” [See Jain 23:67-24:3.]

The three-dimensional information is then extracted using some simplifying assumptions, namely, that the players are in a single plane [Jain 23:50-52], that the camera position is at a fixed known location, and that the rotation angle is zero [Jain 23, 42-44]. This three-dimensional information is then used to determine specific player position in each separate video image, as indicated by a three dimensional cursor, [Jain 21:30-61; and to determine which of the cameras has the best view of a chosen player. [Jain 24:5-11.] The three separate camera images are never combined or even used together. Rather, they are always processed separately.

Separately, in sections 7 through 9 and in figures 12 through 21, Jain describes a completely different system which uses 3 cameras to map the movement in a courtyard in the Engineering School at the University of California, San Diego. [Jain, 25:64-26:4.] A schematic diagram of the cameras and the courtyard is shown in FIG 11a. The system is not stand-alone by any measure, rather, to understand the recorded movement specific steps must be taken. First, the locations and orientation of each camera was calibrated using, among other things, “pre-computed camera coverage tables.” [Jain 30:5-10.] Second, “a complete, geometric three dimensional model of the courtyard was built using map data.” Third, using the map data, the pre-computed camera coverage tables and data interactively received from an image of a user at a known location within the courtyard, the three-dimensional model was calibrated to match the actual courtyard. [Jain 30:14-22.]

Once the cameras are set up, “the three dimensional position of each dynamic object detected by its motion segment component” by using not only a priori information about the

scene and the calibration parameters, but also by assuming that all dynamic objects move on the ground surface. [Jain 28:35-41.] Even more simplifying assumptions are made, in that “pre-computed static mask areas delineating portions of a camera view which cannot have any interesting motion” are not considered. [Jain 32:16-19.] Furthermore, “video processing is limited by focus of attention rectangles.” [See Jain 32: 11-25.] Although the “image to ground projection” shown in FIG. 12 is not explicitly explained in the text, as one of the simplifying assumptions is that “all dynamic objects move on the ground surface” it matches most closely with the following step “the footprint of each bounding box is projected to the primary surface of motion by intersecting a ray drawn from the optic center of that particular camera through the foot of the optic center of that particular camera through the foot of the bounding box with the ground surface.” [Jain 32:62-65.] Thus, the image to ground projection is the process of projecting a portion of a frame as captured by the camera at a known, fixed position onto the two dimensional grid of movement—the ground surface.

B. Lee

The intent of Lee is to provide better techniques to compress transmitted video data by more efficiently encoding motion vectors. This is done by encoding a motion vector for a “feature point”—a series of pixels acting in unison—rather than encoding motion vectors for the individual pixels.

2. Claim Rejection under 35 USC § 112

Claim 36 is objected to because of the term “it” which appears within. The term “it” has been replaced with “said partial model.” This change was made for clarification and not for any reasons related to patentability. As the change requested by the Examiner has been made, Applicants respectfully requests withdrawal of the objection.

3. Claim Rejection under 35 USC § 102

Claim 37 is rejected under 35 U.S.C. 102(b) as being anticipated by USP 5,729,471 to Jain et al. (“Jain”).

Independent claim 37:

Amended claim 37 recites:

An apparatus for recovering a three-dimensional scene from a sequence of two-dimensional frames by segmenting the frames, comprising:

...

means for calculating a partial model for each segment that includes three-dimensional coordinates and camera pose for features within the frames of the segment, the three-dimensional coordinates and camera pose being derived from the frames of the segment;

means for extracting virtual key frames from each partial model.

The applied reference *Jain* fails to teach or suggest many aspects of Applicants’ claim 37. For instance, *Jain* fails to teach or suggest “*calculating a partial model for each segment that includes three-dimensional coordinates and camera pose for features within the frames of the segment, the three-dimensional coordinates and camera pose being derived from the frames of the segment.*”

The Examiner states that the “image to ground projection” of FIG. 12 of *Jain* is “used to calculate and project an image or a partial model for each segment that includes three-dimensional occupancy information...” [Action, pp. 12-13.]

Applicants respectfully disagree. To begin, *Jain* does not teach or suggest “an apparatus for recovering a three-dimensional scene from a sequence of two-dimensional frames.” [Claim 37.] Rather, *Jain* describes “extract[ing] three-dimensional information from video frames captured by cameras.” [Jain 21:64-66.] The extraction of three-dimensional information in *Jain* requires not only the video camera information, but also “pre-computed camera coverage tables,” [Jain 30:5-10] and “a complete, geometric three dimensional model” of the area whose three-dimensional information is being extracted. [Jain 30:14-22.] Applicants have

included the language “the three-dimensional coordinates and camera pose being derived from the frames of the segment” to more clearly point out the differences with the cited art. As within Jain, three-dimensional coordinates are derived from camera coverage tables, a three dimensional model, and the video frames, Jain fails to teach or suggest the partial model having three-dimensional coordinates and camera pose for features within the frames of the segment derived from the frames of the segment.

The Examiner suggests that a combination of the “image to ground projection” shown in FIG. 12, the three-dimensional object estimation also shown in FIG. 12, and the equations that include three dimensional coordinates together teach a “partial model for each segment.” But, Jain does not so teach. As described in section 1A, above, the “image to ground projection” in *Jain*, to the extent it is explained, appears to be a projection of a portion of a video frame which represents an object onto the two dimensional grid of movement—and thus, is a two-dimensional object (the footprint) which only has location values. [Jain 32:2-65.] Further, the three-dimensional occupancy estimation shown in FIG. 12 is, most likely, the process of testing the footprint obtained from the “image to ground projection” to best determine if it matches a footprint of a known object. [Jain, 32:66-33:4.] The “image to ground projection” is not a segment, as it is the projection of one object onto a different plane. Furthermore, as the amended claim language requires that partial model for the segment requires that the feature pose

Jain does indicate that “all supporting observations are used (with appropriate weighting based on distance from the camera, direction of motion, etc.) to update the position of each object.” [Jain 33:3-5.] However, the mere suggestion of updating the position of an object using camera information does not teach or suggest the quoted claim language. Applicants fail to see how the above passage would lead one of ordinary skill in the art to the claimed arrangement, which involves “*means for calculating a partial model for each segment that includes three-dimensional coordinates and camera pose for features within the frames of the segment, the three-dimensional coordinates and camera pose being derived from the frames of the segment;*”

The Examiner also suggests that portions of Jain disclose “the use of equations that includes three dimensional coordinates (x, y, z) that includes camera position or pose camera angle and camera parameter” [Action, p. 13.] It is true that Jain does describe, in section 5.2, using information derived from the video frames along with various matrix calculations to obtain three-dimensional information from the two-dimensional frames. However, the matrix calculations require information apart from that included in the frames themselves. Namely, the cameras are carefully calibrated such that a correlation between points in the world and points in the cameras are not only known, but also used in each matrix calculation, as described in greater detail in section 1A. This teaches directly away from the claim 37 language “the three-dimensional coordinates and camera pose being derived *from the frames of the segment*.”

Since the cited reference fails to describe at least one element recited in claim 37, Applicants request the rejection of claim 37 be withdrawn.

4. Claims Rejections Under 35 USC § 103

Claims 1-2, 4-9, 11-16, and 18-36 are rejected under 35 U.S.C. 103(a) as being unpatentable over USP 5,729,471 *Jain et al.* (“*Jain*”) in view of USP 5,612,743 to Lee (“*Lee*”). Applicants respectfully submit that the claims in their present form are allowable over the applied art. To establish a *prima facie* case of obviousness, three basic criteria must be met. First, there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to modify the reference or to combine reference teachings. Second, there must be a reasonable expectation of success. Finally, the prior art reference (or references when combined) must teach or suggest all the claim limitations. (MPEP § 2142.)

A. The combination of Jain and Lee to reject claims 1-2, 4-9, 11-16, and 18-36 is improper.

The combination of Jain and Lee proposed by the Examiner to reject claims 1-2, 4-9, 11-16, and 18-36 is improper. As to Jain, the Examiner states that “Jain does not specifically

disclose the determining at least a minimum number of feature points being tracked.” [Action, p. 16.] The Applicants agree. The Examiner argues, however, that Lee teaches “the use of threshold values TH and comparison of threshold values of feature points between the current frame the reference frame to check if the threshold is exceeded, thus there is a minimum number of feature points that is determined. Therefore, it would have been obvious to one of ordinary skill in the art to combine the teachings of Jain and Lee, as a whole, for improving the encoding of video image data so as to accurately encode images via the selection of feature points according to the motion of objects in a financially robust manner.” [Action, page 16.]

Applicants respectfully disagree. Even if, for the sake of argument, Jain could be modified as suggested by the Examiner, this is not enough to make the Examiner’s proposed modification obvious. [MPEP 2143.01; see also MPEP 2142.01 and 2145.X.C and D.] In fact, as the frames in Lee are intermediate frames within a video compression stream that must be decoded prior to use, while the frames of Jain are human-viewable video frames that are not compressed, the Examiner’s proposed modification changes the principle of operation of Jain and is thus improper. [See *In re Ratti*, 270 F.2d 810, MPEP § 2143.01.] In addition, Jain and Lee teach away from the combination suggested by the Examiner.

The emphasis of Lee is to provide better techniques to compress transmitted video data by more efficiently encoding the motion vectors within the video data. Temporal redundancies occur between neighboring pixels in different frames. Motion estimation reduces temporal redundancy in successive video frames (interframes) by encoding “motion vectors,” which predict how portions of the frame behave over several frames. Lee teaches encoding a single vector for a “feature point”—a series of pixels acting in unison—rather than encoding separate motion vectors for the individual pixels. The encoded motion vectors are then used to create the compressed video stream. [Lee, 1:15-2:56; FIG. 2.]

Jain, on the other hand has nothing to do with compressing video images. Rather, Jain describes manually processing a series of normal camera picture images to mark the locations of specific football players and specific football field markings. [Jain 22:11-15.]

Thus, the two patents, Jain, and Lee, use similar language for describing entirely

different concepts. For example, the word “frame” in Jain means a picture frame of film that can be viewed and understood without further processing, such as the frames shown in Jain, FIG. 8. However, in Lee, “frame” refers to a frame within a video compression stream, which has been (or will be) encoded, and must be decoded to produce a recognizable video frame.

Furthermore, in Lee, some compressed video frames (reference frames) are encoded without reference to other frames, while other frames, such as predicted current frames, contain the differences with a key frame. Frames that encode motion vectors, are by their nature, predicted current frames, and so must not only be decoded to produce a recognizable image, but must also refer to one or more reference frames during the decoding process to recreate the original image.

The intermediate portion of the compressed stream--the predicted “frame” of Lee, which must be, at a minimum, cross-referenced to another frame and then decoded to be usable, is not equivalent to a standard video frame of Jain, and so cannot be used to substitute for it. Therefore, at least, as Jain describes using standard video frames, which are not an intermediate portion of a compressed video stream, and cannot be substituted for such an intermediate portion, the Examiner’s proposed modification would change this principle of operation of Jain. Therefore, the two references cannot be combined.

Moreover, the phrase “feature point” means two completely different things in the two patents. In Lee, a “feature point” refers to “pixels which are capable of representing the motions of objects in the frame” [Lee 4:58-60] and are automatically extracted by comparing the difference between locations in successive frames and choosing regions whose pixels are different by a threshold amount. That is, locations that move together are determined to be features and are encoded similarly.

In Jain, a “feature point” is the location of a known player, or a known field mark and is marked manually. A person determines, for example, where the player “Washington” is, and marks that location on a video frame. [Jain 22:4-7; 22:23-24.] The feature points of Lee, which are used to efficiently compress similar frames within a compressed video stream, have nothing to do, other than language overlap, with the feature points of Jain, which are hand-

selected player and field locations. For this reason as well, the two references cannot be combined. Furthermore, as Jain describes picking out feature points by having a human mark the feet of a football player within a video still, while Lee only discusses the relative movement of pixel in successive frames, the Examiner's proposed modification would change this principle of operation of Jain.

In addition, the motivation the Examiner cites to modify Jain with Lee is improper. The Examiner writes that the motivation to modify Jain with the teachings of Lee is that "it would have been obvious to one of ordinary skill in the art to combine the teachings of Jain and Lee, as a whole, for improving the encoding of video image data so as to accurately encode images via the selection of feature points according to the motion of objects in a financially robust manner." However, the modification suggested by the Examiner would, even if possible, produce no such result, as the feature points in Lee are used to effectively compress and encode the image, and so will be decompressed and decoded by the time the image is reconstructed into traditional video frames, such as those used by Jain. Thus, using Lee will have no effect at all on the feature point selection of Jain.

For at least these reasons, as Jain and Lee are not properly combined, claims 1-2, 4-9, 11-16, and 18-36 should be allowable.

B. Jain and Lee, either in combination or separately, fail to teach at least one element of claims 1-2, 4-9, 11-16, and 18-36.

Independent Claim 1

Amended claim 1 recites as follows:

dividing the sequence of frames into frame segments wherein the frames in the sequence comprise feature points and wherein the sequence of frames is divided into frame segments based upon frames in each frame segment having at least a minimum number of feature points being tracked to at least one base frame in the frame segment...

The applied references, *Jain* and *Lee*, both individually and in combination, fail to teach or suggest many aspects of claim 1. For instance, combining *Jain*'s capturing video shot sequences at a standard 30 frames per second with *Lee*'s comparing "a current frame" to "a reference frame" for determining pixel differential values for various regions still fails to teach or suggest *"dividing the sequence of frames into frame segments wherein the frames in the sequence comprise feature points and wherein the sequence of frames is divided into frame segments based upon frames in each frame segment having at least a minimum number of feature points being tracked to at least one base frame in the frame segment."*

The Action relies on *Jain* and *Lee*. First of all, as the Action agrees, "*Jain* does not specifically disclose determining whether a threshold number of feature points being tracked." [See, Action at p. 16, ll. 8-9.] Instead, the Action relies on *Lee* comparing a "current frame" (pixel-by-pixel) to a "reference frame" to determine a "differential pixel value" and further processing including "the comparison block 313 compares on a pixel-by-pixel basis each of the differential pixel values included in the difference signal with a threshold value TH...if a differential pixel value is less than the threshold value TH, it is set to the conversion value 0. Otherwise, the differential pixel value is set to the conversion value 1." [See *Lee* 5: 22-32.]

Lee does not use the threshold to track the number of feature points. Rather, the threshold is used to partition a frame into two discontinuous regions equal to 0 and 1. If "the differential pixel value is less than then threshold value TH, it is set to the conversion value 0. Otherwise, it is set to the conversion value 1." [Lee 5:28-32] Such a partitioned frame is shown in FIG. 4. "There are two distinct zones in the error frame 41: one is the regions (e.g., A B and C) with the conversion value 1; and the other, with the conversion value 0." [Lee 5:35-37.] Feature points are never counted in *Lee*, let alone tracked. Further, as the feature points are not tracked, they are not tracked to a base frame. What *Lee* does teach, partitioning a region into two zones, does not teach or suggest *"frames in each frame segment having at least a minimum number of feature points being tracked to at least one base frame in the frame segment."*

Since the applied references, do not teach or suggest at least one element of claim 1, claim 1 in its present form should be allowed.

Dependent claims 2 and 4-8

Claims 2 and 4-8, ultimately depend on claim 1 and, thus, at least for the reasons set forth above with respect to claim 1, claims 2 and 4-8 should be also be allowed. Furthermore, each of the claims 2 and 4-8 also recites independently patentable features and, thus, should be allowed for that reason.

Independent claim 9

Claim 9 recites as follows:

A method of recovering a three-dimensional scene from two-dimensional images, the method comprising:

...

dividing the sequence of frames into segments, ...

for each segment, encoding the frames in the segment into at least two virtual frames that include a three-dimensional structure for the segment and an uncertainty associated with the segment

for each of the at least two chosen frames, projecting a plurality of three-dimensional points into a corresponding virtual frame; and

for each of the at least two chosen frames, projecting an uncertainty into the corresponding virtual frame.

The applied reference Jain fails to teach or suggest many aspects of Applicants' claim 9. For instance, Jain fails to teach or suggest "*a method of recovering a three-dimensional scene from two-dimensional images, the method comprising... dividing the sequence of frames into segments, ... for each segment, encoding the frames in the segment into at least two virtual frames that include a three-dimensional structure for the segment and an uncertainty associated with the segment...*for each of the at least two chosen frames, projecting a plurality of three-dimensional points into a corresponding virtual frame; and for each of the at least two chosen frames, projecting an uncertainty into the corresponding virtual frame" as recited in Applicants' claim 9.

The Examiner relies on the following passage to teach “*at least two virtual frames that include a three-dimensional structure for the segment and an uncertainty associated with the segment... for each segment, encoding the frames in the segment into at least two virtual frames that include a three-dimensional structure for the segment and an uncertainty associated with the segment....*”

“Accordingly, estimates had to be made for those video frames that didn’t show enough obvious known points. The results of such estimations are not necessarily accurate. Many known points an (sic) this image can be used for camera calibrations.” [Jain 24:63-67.] In Jain, as described with reference to Section 1A, humans mark the location of a player’s feet and any field markings on selected video frames. [See Jain 22:11-15.] If the feet of a player are not visible, and/or there are no visible field markings, as shown with reference to FIG. 9b, feature points (also called ‘known points’) cannot be accurately placed. Therefore, the locations of such feature points were estimated. Thus, a human would mark the best guess as to where, say, a field marking may be located on a specific frame. [See Jain 24:51-67.] Thus, Jain acknowledges that the specific protocol used to determine feature points leads to some unknown amount of inaccuracy. There is no determination in *Jain*, however, as to which frames may have errors, how many frames contain errors, or what the magnitude of such errors may be. Furthermore, not only is there no quantification of the error rate, but there is no procedure, process, or method disclosed in *Jain* which could use the error information in a productive fashion. Thus, Jain leads away from, a minimum, calculating an uncertainty associated with the segment.

As there is no uncertainty calculated for any given segment, then Jain **cannot** include such an uncertainty associated with the segment in a virtual frame.

Furthermore, neither Jain nor Lee teach or suggest the claim language “*at least two virtual frames that include a three-dimensional structure for the segment.*” The Examiner states that “segmented frames are encoded into at least two virtual key frames **to ascertain** the best, possible three-dimensional reconstruction of the two-dimensional frame data to yield the 3D visualization. [See Action, p. 20.] However, ascertaining is not the same as “including.”

There is no teaching or suggestion in Jain that the key frames of Jain include a three-dimensional structure for the segment.

Thus, Jain fails to teach or suggest several elements of claim 9. The reference Lee, either separately or in combination with Jain, also fails to teach or suggest the language of claim 9. Since the cited references do not teach or suggest at least one element of claim 9, claim 9 in its present form should be allowed.

Dependent claims 10-22

Claims 10-22, ultimately depend on claim 9 and thus, at least for the reasons set forth above with respect to claim 9, claims 10-22 should be also be allowed. Furthermore, each of the claims 10-22 also recites independently patentable features and, thus, should be allowed for that reason.

Independent Claim 23

The applied references Jain and Lee fail to teach or suggest many aspects of Applicants' claim 23. For instance, Jain and Lee both fail to teach or suggest

“(e) determining a number of the selected feature points from the base frame that are also identified in the next frame; and

(f) if the number of the selected feature points from the base frame that are also identified in the next frame is greater than or equal to a threshold number, adding the next frame to the first segment of frames of the sequence.”

The examiner states: “Jain does not specifically disclose the adding the second frame to the segment.” [Action, p. 25, l. 15.] Applicants respectfully agree. Applicants respectfully suggest that, at a minimum, an obviousness rejection should include a reference where the cited language is taught. Since the cited references do not teach or suggest at least the cited portions of claim 23, Applicants respectfully suggest that this claim is in condition for allowance.

Furthermore, it is impermissible to use the claims as an instruction manual or

“template” to piece together the teachings of the prior art to render a claimed invention obvious. *See In re Fritch*, 972 F.2d 1260, 1266. “Virtually all [inventions] are combinations of old elements.” *In re Rouffet*, 47 USPQ.2d at 1457 (Fed Cir. 1998). Thus, the invention “must be viewed not after the blueprint has been drawn by the inventor, but as it would have been perceived in the state of the art that existed at the time the invention was made without hindsight or knowledge of the invention.” *Sensonics, Inc. v. Aerosonic Corp.*, 38 USPQ.2d 1551, 1554 (Fed. Cir. 1996). “To draw on hindsight knowledge of the patented invention, when the prior art does not contain or suggest that knowledge, is to use the invention as a template for its own reconstruction -- an illogical and inappropriate process by which to determine patentability” *Id.*

The best argument against “hindsight-based obviousness analysis is rigorous application of the requirement for a showing of the teaching or motivation to combine prior art references.” *In Re Dembicziak*, 50 USPQ.2d 1614, 1617 (Fed. Cir. 1999).

For the reasons stated above, the general statement in the Office Action to combine Jain and with the teachings in the application “for accurately enhancing the three-dimensional representation of the targeted scene” is legally insufficient in that the office has failed to prove that they have not used hindsight to combine the teachings of Jain with the teachings found in the claims themselves. [Action, at p. 29, ll. 1-5.] The suspicion that hindsight was used is reinforced by the fact that no reference was provided for the claim language “adding the next frame to the first segment of frames of the sequence” as recited in claim 23.

As impermissible hindsight was used to reconstruct claim 23, this claim is in condition for allowance.

In addition, Jain does not disclose the modification proposed by the examiner; i.e., the manual adjustment of the number of key frames. The Examiner continues: “However Jain discloses the manual adjustment of the number of key frames, where the number is one key frame for every thirty frames, i.e., a segment.” [Action, p. 25, ll. 26-27.] Applicants respectfully disagree. Jain discloses manually **selecting** a key frame in each group of 30 frames, as shown in the following quote: “Therefore, one key frame has been manually

selected for every thirty frames, and scene analysis has been applied to the selected key frames.” [Jain 23:64-67.]

However, selecting a frame is different than adjusting the number of key frames. Nowhere does Jain suggest “adjusting” or changing the **number** of frames from which a key frame is chosen. Further, 30 frames is the standard NTSC frame rate for a second of film. [Action, p. 3, l. 22.] In Jain, a key frame is selected for each second (30 frames) of film. The number 30 was not chosen randomly, which leads away from adjusting or changing the number of frames (30) from which a key frame is chosen.

Moreover, Jain actively teaches against the modification claimed by the examiner, as the only frames that are considered within each video sequence in Jain are the key frames.

Specifically, for each video sequence, key frames are selected manually for every thirty frames. [Jain 23:64-65.] For each selected key frame, a person manually selects feature points. [Jain 22:11-15.] Then, the locations of the marked feature points within the key frames are interpolated—essentially discarding the other 29 frames, as shown by the following passage: “For frames in between, player position and camera status is estimated by interpolation between key frames by proceeding under the assumption that coordinate values change linearly between a consecutive two key frames.” [Jain 23:67-24:3.] Thus, as all the frames between the key frames are discarded, Jain teaches against using those discarded frames for anything, let alone “adding the next frame to the first segment of frames of the sequence” as taught in claim 23. The reference Lee, either separately or in combination with Jain, also fails to teach or suggest the language of claim 23.

For all of the reasons mentioned above, claim 23 is in condition for allowance.

Since the applied references do not teach or suggest at least one element of claim 23, claim 23 in its present form should be allowed.

Dependent claims 24-30

Claims 24-30 depend on claim 23 and thus, at least for the reasons set forth above with respect to claim 23, claims 24-30 should be allowed. Furthermore, each of the claims 24-30 also recites independently patentable features and, thus, should be allowed for that reason.

Independent claim 31

Claim 31 recites as follows:

dividing a long sequence of frames into segments and reducing the number of frames in each segment by representing the segments using between two and five representative frames per segment....

The applied references Jain and Lee fail to teach or suggest many aspects of Applicants' claim 31. For instance, Jain and Lee both fail to teach or suggest "dividing a long sequence of frames into segments and reducing the number of frames in each segment by representing the segments using between two and five representative frames per segment."

The Examiner states that "Jain does not specifically disclose the reducing the number of frames in each segment by representing the segments using between two and five representative frames per segment." [Action at p. 28, lines 18-20.] Applicants agree. The Examiner then states: "However, Jain discloses the manual adjustment of the number of key frames, where the number is one key frame for every thirty frames, i.e., a segment. Therefore, since Jain teaches the manual adjustment of one key frame or representative frame for every thirty frames, it would have been obvious to one of ordinary skill in the art to manually change the number of key (representative) frames per segment from anywhere between two to five key or representative frames per segment if necessary for accurately enhancing the three-dimensional representation of the targeted scene." [Action at p. 28, lines 18-20.] Applicants respectfully disagree.

The Examiner has not provided a reference which teaches "reducing the number of frames in each segment by representing the segments using between two and five representative frames per segment" as recited in claim 31. Applicants respectfully suggest

that, at a minimum, an obviousness rejection should include a reference where the cited language is taught. Since the cited references do not teach or suggest at least the cited portions of claim 31, Applicants respectfully suggest that this claim is in condition for allowance.

Furthermore, it is impermissible to use the claims as an instruction manual or “template” to piece together the teachings of the prior art to render a claimed invention obvious. *See In re Fritch*, 972 F.2d 1260, 1266. “Virtually all [inventions] are combinations of old elements.” *In re Rouffet*, 47 USPQ.2d at 1457 (Fed Cir. 1998). Thus, the invention “must be viewed not after the blueprint has been drawn by the inventor, but as it would have been perceived in the state of the art that existed at the time the invention was made without hindsight or knowledge of the invention.” *Sensonics, Inc. v. Aerosonic Corp.*, 38 USPQ.2d 1551, 1554 (Fed. Cir. 1996). “To draw on hindsight knowledge of the patented invention, when the prior art does not contain or suggest that knowledge, is to use the invention as a template for its own reconstruction -- an illogical and inappropriate process by which to determine patentability” *Id.*

The best argument against “hindsight-based obviousness analysis is rigorous application of the requirement for a showing of the teaching or motivation to combine prior art references.” *In Re Dembicziak*, 50 USPQ.2d 1614, 1617 (Fed. Cir. 1999).

For the reasons stated above, the general statement in the Office Action to combine Jain and with the teachings in the application “for accurately enhancing the three-dimensional representation of the targeted scene” is legally insufficient in that the office has failed to prove that they have not used hindsight to combine the teachings of Jain with the teachings found in the claims themselves. [Action, at p. 29, ll. 1-5.] The suspicion that hindsight was used is reinforced by the fact that no reference was provided for the claim language “reducing the number of frames in each segment by representing the segments using between two and five representative frames per segment” as recited in claim 31.

As impermissible hindsight was used to reconstruct claim 31, this claim is in condition for allowance.

Furthermore, Jain actively teaches against the modification claimed by the examiner, as

the only frames that are considered within each video sequence in Jain are the key frames.

Specifically, for each video sequence, key frames are selected manually for every thirty frames. [Jain 23:64-65.] For each selected key frame, a person manually selects feature points. [Jain 22:11-15.] Then, the locations of the marked feature points within the key frames are interpolated—essentially discarding the other 29 frames, as shown by the following passage: “For frames in between, player position and camera status is estimated by interpolation between key frames by proceeding under the assumption that coordinate values change linearly between a consecutive two key frames.” [See Jain 23:67-24:3.] Thus, as all the frames between the key frames are discarded, Jain teaches against using those discarded frames for anything, let alone “representing the segments using between two and five representative frames per segment” as taught in claim 31. Moreover, there is no teaching or suggestion in Jain that the particular range of numbers cited, 2-5, should be used, or that such a number of frames would offer any benefit. The reference Lee, either separately or in combination with Jain, also fails to teach or suggest the language of claim 31.

For all of the reasons mentioned above, claim 31 is in condition for allowance.

Dependent claims 32 - 35

Claims 32 -35 depend on claim 31 and, thus, at least for the reasons set forth above with respect to claim 31, claims 32 - 35 should be also be allowed. Furthermore, each of the claims 32 - 35 also recites independently patentable features and, thus, should be allowed for that reason.

Independent claim 36

Claim 36 recites as follows:

extracting virtual key frames from each partial model, the virtual key frames having three-dimensional coordinates for the frames and an uncertainty associated with the frames....

The applied reference Jain fails to teach or suggest many aspects of Applicants’ claim

36. For instance, Jain fails to teach or suggest “extracting virtual key frames from each partial model, the virtual key frames having three-dimensional coordinates for the frames and an uncertainty associated with the frames.” The Examiner relies on the following passage to teach “*extracting virtual key frames from each partial model, the virtual key frames having three-dimensional coordinates for the frames and an uncertainty associated with the frames.*”

“Accordingly, estimates had to be made for those video frames that didn’t show enough obvious known points. The results of such estimations are not necessarily accurate. Many known points an (sic) this image can be used for camera calibrations.” [Jain 24:63-67.] In Jain, as described with reference to Section 1A, humans mark the location of a player’s feet and any field markings on selected video frames. [See Jain 22:11-15.] If the feet of a player are not visible, and/or there are no visible field markings, as shown with reference to FIG. 9b, feature points (also called ‘known points’) cannot be accurately placed. Therefore, the locations of such feature points were estimated. Thus, a human would mark the best guess as to where, say, a field marking may be located on a specific frame. [See Jain 24:51-67.] Thus, Jain acknowledges that the specific protocol used to determine feature points leads to some unknown amount of inaccuracy. There is no determination in *Jain*, however, as to which frames may have errors, how many frames contain errors, or what the magnitude of such errors may be. Furthermore, not only is there no quantification of the error rate, but there is no procedure, process, or method disclosed in *Jain* which could use the error information in a productive fashion. Thus, Jain leads away from calculating an uncertainty associated with the segment.

As there is no uncertainty calculated for any given segment, then Jain **cannot** include such an uncertainty associated with the segment in a virtual frame.

The reference to Lee, either separately or in combination with Jain, also fails to teach or suggest the language of claim 36.

Thus, as Jain and Lee fail to teach or suggest at least one element of claim 36, claim 36 in its present form should be allowed.

Conclusion

The claims in their present form should now be allowable. Such action is respectfully requested.

Respectfully submitted,

KLARQUIST SPARKMAN, LLP

One World Trade Center, Suite 1600
121 S.W. Salmon Street
Portland, Oregon 97204
Telephone: (503) 595-5300
Facsimile: (503) 595-5301

By /Genie Lyons/
Genie Lyons
Registration No. 43,841